Kongeriget Danmark

Patent application No.: PA 2002 00591

Date of filing: 19 April 2002

Applicant: (Name and address)
BIOIMAGE A/S
Mørkhøj Bygade 28
DK-2860 Søborg
Denmark

Title: Screening using split GFP

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

9 May 2003

Åse Damm

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

PATENT- OG VAREMÆRKESTYRELSEN

Field of invention

The present invention relates to various uses of split fluorophore complementation in relation to *inter alia* screening for drugs capable of modulating protein-protein interactions, screening for participants in specific protein-protein interactions, screening for protein-protein interactions modulated by specific compounds.

Background of the invention

It has been suggested to use the reassembly of certain enzyme fragments to the complete enzyme as a measure of protein-protein interactions. PNAS 91, 1994 - Johnsson discloses reassembly of Ubiquitin. This reassembly is detected through the irreversible cleavage of the fusion by Ubiquitin protease and release of a Reporter. As opposed to the two-hybrid technique, this technique includes the possibility of monitoring a protein-protein interaction as a function of time, at the natural sites of this interaction in a living cell.

A similar system is suggested for the reassembly of β-galactosidase (PNAS 94, 1997 - Rossi), DHFR (WO98/34120) and Green Fluorescent Protein (GFP) (J.Am.Chem.Soc 122, 2000 - Ghosh and WO01/87919). The basic concept is that by splitting a functional protein in two fragments, the function is lost. The two fragments are then transfected into cells fused in frame to proteins X and Y, respectively. Binding between proteins X and Y will bring the two fragments so close together that the functional protein will regain its function. If the function is DHFR, the cells will survive only if proteins X and Y bind to each other. If the function is fluorescence, the cells will emit light upon excitation only if protein X and protein Y bind to each other.

The present invention describes improvements and new uses of this reassembly technique, the technique of fluorescence complementation.

Detailed disclosure

When the word "fluorophore" is used in the present application, it is meant to indicate a fluorescent protein, that is a protein that, when expressed by a cell, emits fluorescence upon exposure to light of the correct excitation wavelength (e.g. as described by Chalfie,

M. *et al.* (1994) Science 263, 802-805). One or more amino acids of the fluorophore may have been substituted, inserted or deleted. Fluorophore, as used herein, includes *inter alia* wild-type Green Fluorescent Protein (GFP) derived from the jelly fish *Aequorea Victoria* , or from other members of the Coelenterata, such as the red fluorescent protein from *Discosoma* sp. (Matz, M.V. *et al.* 1999, Nature Biotechnology 17: 969-973), GFP from *Renilla reniformis*, GFP from *Renilla Muelleri* or fluorescent proteins from other animals, fungi or plants. The term also includes modifications of GFP, such as the blue fluorescent variant of GFP disclosed by Heim et al. (Heim, R. *et al.*, 1994, Proc.Natl.Acad.Sci. 91:26, pp 12501-12504), and other modifications that change the spectral properties of the GFP fluorescence, or modifications that exhibit increased fluorescence when expressed in cells at a temperature above about 30°C described in PCT/DK96/00051, published as WO 97/11094 on 27 March 1997, and that comprises a fluorescent protein derived from *Aequorea* Green Fluorescent Protein or any functional analogue thereof, wherein the amino acid in position 1 upstream from the chromophore has been mutated to provide an increase of fluorescence intensity when the fluorescent protein of the invention is expressed in cells.

Numerous cell systems for transfection exist. A few examples are Xenopus oocytes or insect cells, such as the sf9 cell line, or mammalian cells isolated directly from tissues or organs taken from healthy or diseased animals (primary cells), or transformed mammalian cells capable of indefinite replication under cell culture conditions (cell lines). However, it is preferred that the cells used are mammalian cells. This is due to the complex biochemical interactions specific for each cell type. The term "mammalian cell" is intended to indicate any living cell of mammalian origin. The cell may be an established cell line, many of which are available from The American Type Culture Collection (ATCC, Virginia, USA) or similar Cell Culture Collections. The cell may be a primary cell with a limited life span derived from a mammalian tissue, including tissues derived from a transgenic animal, or a newly established immortal cell line derived from a mammalian tissue including transgenic tissues, or a hybrid cell or cell line derived by fusing different cell types of mammalian origin e.g. hybridoma cell lines. The cells may optionally express one or more non-native gene products, e.g. receptors, enzymes, enzyme substrates, prior to or in addition to the fluorescent probe. Preferred cell lines include, but are not limited to, those of fibroblast origin, e.g. BHK, CHO, BALB, NIH-3T3 or of endothelial origin, e.g. HUVEC, BAE (bovine artery endothelial), CPAE (cow pulmonary artery endothelial), HLMVEC (human lung micro vascular endothelial cells), or of airway epithelial origin, e.g.

BEAS-2B, or of pancreatic origin, e.g. RIN, INS-1, MIN6, bTC3, aTC6, bTC6, HIT, or of hematopoietic origin, e.g. primary isolated human monocytes, macrophages, neutrophils, basophils, eosinophils and lymphocyte populations, AML-14, AML-193, HL-60, RBL-1, U937, RAW, JAWS, or of adipocyte origin, e.g. 3T3-L1, human pre-adipocytes, or of neuroendocrine origin, e.g. AtT20, PC12, GH3, muscle origin, e.g. SKMC, A10, C2C12, renal origin, e.g. HEK 293, LLC-PK1, or of neuronal origin, e.g. SK-N-DZ, SK-N-BE(2), HCN-1A, NT2/D1.

The non-fluorescent fragments of fluorescent proteins that can be combined to form one functional fluorescent unit are usually produced by splitting the coding nucleotide sequence of one fluorescent protein at an appropriate site and expressing each nucleotide sequence fragment independently. The fluorescent protein fragments may be expressed alone or in fusion with one or more protein fusion partners. Each translated sequence must contain appropriate start and stop codons and some residues in the fluorescent protein may be encoded by both coding sequences whereas other residues may not be encoded by any of the coding sequences.

We have data on hand suggesting to split EGFP at amino acid 157-158, 172-173, or at amino acid144-145. Based on these findings it is concluded that appropriate splitting sites in GFP are located in the loop regions between the residues that form the beta-sheet structures of the GFP beta-barrel. Accordingly, splits in GFP must be made next to or close to residues 23, 39, 51, 102, 116, 157, or 173 or within, next to or close to the regions defined by residues 76-90, 129-144, 189-197, or 209-215 (Fig. 1). All residues are numbered according to the numbering of wild type A. victoria GFP (Genbank accession no. M62653) and said numbering also applies to equivalent positions in homologous sequences exemplified by alignment #1 of fluorescent protein sequences.

The choice of split site for a particular assay, depends on the properties needed for the fluorophore as it is presently assumed that the various split sites will have various influences on the different speed of folding, different intensity etc of the maturing or matured fluorophore.

Thus, one aspect of the invention relates to a method for generating a library of interacting proteins within living cells consisting of:

1. Introducing into a pool of cells two sets of plasmids, either simultaneously or sequentially, one set of plasmids encoding a library of proteins A each fused to the N-terminal half of the complementing fluorophore and the second set of plasmids encoding a library of proteins B each fused to the C-terminal half of the complementing fluorophore.

2. Sorting the cells into those where a functional fluorophore has been formed and those where a functional fluorophore has not been formed, the formation of said functional fluorophore being indicative of an interaction having occurred between proteins A and B within the cell.

In a preferred embodiment of the invention, the sorting in step 2 is done by FACS.

Another aspect of the invention relates to a method for assessing the general utility of a compound in modulating protein-protein interactions consisting of:

1. Contacting a library of interacting proteins A and B within living cells with the compound.

2. Sorting the cells into those where the fluorophore has been disrupted and those where the fluorophore is intact, the disruption of said fluorophore being indicative of the compound having disrupted the interaction between proteins A and B.

3. Calculating the fraction of interaction pairs A-B that has been split by the compound by calculating the fraction of cells where disruption of the fluorophore has occurred upon contact with the compound.

In a preferred embodiment of the invention, the sorting in step 2 is done by FACS.

Yet another aspect of the invention relates to a method for determining the specific protein-protein interactions inhibited by a compound comprising:

1. Contacting a library of interacting proteins A and B within living cells with the compound.

2. Sorting the cells into those where the fluorophore has been disrupted and those where the fluorophore is intact, the disruption of said fluorophore being indicative of the compound having disrupted the interaction between proteins A and B.

3. Determining the identity of interacting pairs A-B split by the compound.

In a preferred embodiment of the invention, the sorting in step 2 is done by FACS.

As described in e.g. WO98/45704, the measurement of translocation of protein X from one site to another reveals crucial information of the cellular dynamics. One aspect of the invention relates to an assay for measuring translocation of protein X comprising the steps of:

1. Transfecting into a cell two constructs, the first construct comprising the sequence encoding an anchor fused to a zipper sequence fused to the first half of the fluorophore, the second construct comprising protein X fused to a zipper sequence fused to the second half of the fluorophore.

2. Induce translocation of X to the anchor-site.

3. Monitor the increase in fluorescence caused when the zippers bring the two halves of the fluorophore in close apposition, and the fluorophore emits light upon excitation.

An example of such use is Split fluorophore complementation as a redistribution sensor: here one cell line that stably expresses Histone-zipper-GFPa fusion is made. Into that cell line the nuclear translocator ie p65-zipper-GFPb is transfected. Upon translocation of p65 into the nucleus the two zipper sequences will cause GFPa and GFPb to fold and mature and be fluorescent.

The present invention includes as anchor components for the method any and all genetically encodable cellular components that have a defined cellular distribution.

Anchor systems can be designed to achieve redistribution to compartments or locations within cells where the proteins of interest will experience the influences that would normally be required to modulate the interaction between those proteins. As an example, some proteins normally require to be phosphorylated or dephosphorylated by enzymes

sequestered in the plane of the plasma membrane – for such proteins of interest it is appropriate to choose an anchor component that would be expected to be confined to the plasma membrane, to allow the interacting proteins to be appropriately modified. Thus, in one embodiment, a preferred anchor component that will target the anchor conjugate to the plasma membrane is a protein containing the transmembrane domain of the epidermal growth factor receptor (EGFR), or containing the transmembrane domain of a protein from the integrin protein family, or containing the myristoylation sequence from c-Src (residues 1-14).

In another embodiment, a histone protein is used as the anchor, or a protein normally restricted to nucleoli, for example the p120 nucleolar protein, in order to direct the anchoring conjugate to the nucleus.

In another embodiment, the anchor protein is chosen from those proteins normally confined to mitochondrial outer or inner membranes for example VDAC, F_o subunit of ATP-ase, or NADH dehydrogenase. In another embodiment, the anchor protein is chosen from the group of proteins normally confined to the various different regions of Golgi bodies for example TGN38 or ADAM12-L. In another embodiment, the anchor protein is chosen from the group of proteins normally confined to focal adhesion complexes for example P125, FAK, integerin alpha or beta, or paxillin. In another embodiment, the anchor protein is chosen from the group of proteins normally associated with cytoskeletal structures such as F-actin strands or micro tubular bundles for example MAP4, kinesins, myosins or dyniens.

One aspect of the invention relates to multiplexing split fluorophore complementation using different colours. By combining one fluorescent protein fragment with two or more appropriate complementary fragments, it is possible to determine the extent of binding of the first fluorescent protein fragment to either of the two other complementary fragments if the two possible fluorescent complexes have distinct fluorescence excitation or emission characteristics or both. Typically, the first fluorescent protein fragment will be fused to a protein that may bind to either of three other proteins each of them being appropriately fused to distinct complementary fragments. For example, the first fluorescent protein fragment can be a C-terminal fragment of enhanced GFP (EGFP, SEQ ID NO: 1) obtained by splitting EGFP after residue 80. The three complementary fragments can be appropriate N-terminal fragments of EGFP, of EGFP Y66W (SEQ ID NO: 2) and EGFP

Y66H (SEQ ID NO: 3), respectively. The three EGFP variants have different spectral characteristics:

Fluorescent protein	Excitation max (nm)	Emission max (nm)
GFP	396	508
GFP Y66W	382	448
GFP Y66H	458	480

Reference: Heim, R., Prasher, D.C., and Tsien, R.Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci .U. S. A. 91, 12501-12504.

For example, fluorescent complexes produced by assembling a C-terminal half of GFP (e.g. residues 158 to 238) with corresponding N-terminal halves (e.g. residues 1-157) of GFP, GFP Y66W, or GFP Y66H will have clearly distinct fluorescence characteristics and the relative amounts complexes in mixtures can be calculated.

Not only various colours can be used as exemplified above, but other physical parameters of the fluorophore can be altered e.g. intensity, fluorescent life-time, folding time etc.

Intracellular signalling is a highly dynamic process where the signalling proteins translocate and undergo reversible interactions with different binding partners. It is likely that several of the cytosolic translocations occur by passive diffusion, and is controlled by the conditional change of affinities between the specific signalling protein and its binding partners (Terusal, M.N. & Meyer, T. Cell 104, 181 (2000)). This system is in a dynamic equilibrium where only minor changes may lead to net movement of the signalling protein towards another binding partner. To study these phenomena in a physiological meaningful way, they need to be performed in a cellular environment. In the present invention, we have expressed multiple binding partners fused to half a fluorophore together with the signalling protein fused to the other half of the fluorophore. The binding partners will be fused to parts of the fluorophore containing different mutations that upon complementation with the other half of the fluorophore (fused to the signalling protein) will exert different spectral properties (e.g. excitation, emission, fluorescence life time a.o.). This will allow quantification of how the signalling protein is divided between the different binding

partners. Also, the change of the equilibrium can be accessed, caused e.g. by a drug or putative drug.

An example of this is the MAPK signalling pathways relying on sequential binding event between multiple enzymes. Taking the Erk pathway: Ras will bind to and activate MEKK1, which will bind to and activate MEK1, which will bind to and activate Erk1. Fusing MEKK1 and Erk1 with two halves of a fluorophore carrying different mutations that upon binding to MEK1 that is fused to the other part of the fluorophore will allow monitoring of the equilibrium between MEKK1:MEK1 binding versus MEK1:Erk1 binding by spectral analysis of the fluorescence from a cell population. Additional information on the specific localisation of the binding events is obtained by performing the detection in a microscope system with spectral analysis capabilities.

The two colour multiplexing has several uses: In most cases, as described above, the protein of interest is linked to the "constant" half of the fluorophore whereas its interaction partners each are linked to "variable" parts of the fluorophore e.g. one that upon fusion gives rise to a green fluorophore and one that gives rise to a blue fluorophore. This can give spatial information, ie the two different interaction partners are in different locations so colour will tell you where your protein of interest is. The interaction partners could also be in the same location so here colour gives you an indication of which interaction partner your protein prefers at any given time. Finally, as a special case of the latter, if your protein's interaction with the different partners is modulated by eg posttranslational modifications, colour can tell you whether your protein is modified or not. These three different setups can be used either as sensors for the physical state of your protein in the broadest possible sense, or they can be used as screening assays where you measure the ability of the test compounds to alter the ratio between the two colour readouts. Finally, colour is only one physical parameter of GFPs. Other physical parameters that can be localised to specific amino acids in the GFP sequence and that are easily detectable, such as absorption spectra, fluorescence lifetime, time for fluorophore maturation etc., could be employed in exactly the same way as colour. The number of different interaction partners need not be limited to two.

One special example shows how this method may be employed as a tool to use spatial information for screening, even if nothing is known about the interaction partners of a particular protein in different compartments (or if it has no interaction partners in one or more of those compartments). Cystic Fibrosis is perhaps the most frequent and well-

studied protein trafficking disease. The cystic fibrosis transmembrane conductance regulator (CFTR) is a multi-membrane spanning protein that normally functions at the apical plasma membrane of airway epithelial cells as a Cl-efflux channel. The most common mutation, (DeltaF508) causes the protein to be retained in the endoplasmic reticulum (ER) and so reduces the amount of CFTR expressed in the plasma membrane of epithelial cells, resulting in decreased Cl- efflux from the cells. It would appear from numerous studies that this ER retention defect of DeltaF508 is reversible, and reduced temperature, some small molecules, and induction of "chaperones" allow DeltaF508 to traffic to the plasma membrane and increase Cl- permeability. One way of screening for compounds that modify mutant CFTR behaviour employs the split fluorophore /multiple colours concept. You could express the mutant CFTR in cells as a fusion with a zipper fragment fused to the "constant" half of the fluorophore. Expression in the same cell fusions of ER and plasma membrane markers with zippers fused to different colours will reveal when the CFTR mutant reaches the ER, this gives rise to one colour. If CFTR is moved to the PM e.g. by drug stimulus, that would give rise to a different colour. So by screening for compounds that favours the generation of the "PM colour", you could find molecules that specifically correct the ER retention defect of CF patients.

One aspect of the invention thus relates to a method for generating a library of interacting proteins within living cells consisting of:

1. Introducing into a pool of cells two sets of plasmids, either simultaneously or sequentially, one set of plasmids encoding a library of proteins A each fused to the N-terminal half of the complementing fluorophore and the second set of plasmids encoding a library of proteins B each fused to the C-terminal half of the complementing fluorophore.

2. Sorting the cells by imaging into those where a functional fluorophore has been formed in the right compartment and those where a functional fluorophore has not been formed or those where a functional fluorophore has been formed but in the wrong cellular compartment, the formation of said functional fluorophore being indicative of an interaction having occurred between proteins A and B within the right cellular compartment.

For example, for a protein moving from one compartment to another compartment in response to a stimuli, the binding partners in each compartment can be identified.

In one embodiment of the invention, the method relates to identification of drugs that will cause disruption of binding between two proteins when located in one cellular compartment but not in another cellular compartment. This embodiment is carried out essentially as described above with the only difference that instead of sorting the cells based on intensity, the cells are imaged with standard imaging equipment to determine not only if binding has taken place, but also where such binding has occurred.

This system is also useful for screening for fluorophores with novel properties such as those that can be used in the split fluorophore complemensation usages described above. By mutagenisis of both the N- and C-termini of GFP in this system, the system is used to screen in a combinatorial manner for double (or more) mutants of the fluorophore with novel properties. This gives a wider selection to choose from. Furthermore, as the two mutations are in different halves of the molecule, they could be additive or compensatory, or both. Finally, the fact that they have been found using the split fluorophore complementation system immediately means that they can be used as sensors in this system.

Based on CaM and M13 each fused to a half fluorophore, the amount of Ca can be determined as fluorescence. This principle can be applied to other systems, where the presence of one component will be reflected by the binding of two proteins (divalent metal ions are known for doing this).

Examples

Example 1: Alignment of fluorescent proteins

Genbank entry	Fluorescent protein
P42212	Aequorea victoria green-fluorescent protein
AF372525	Renilla reniformis green fluorescent protein
AY015996	Renilla muelleri green fluorescent protein
AY013824	Aequorea macrodactyla isolate GFPxm
AF384683	Montastraea cavernosa green fluorescent protein
AF401282	Montastraea faveolata green fluorescent protein
AY015995	Ptilosarcus sp. CSG-2001 green fluorescent protein
AF322221	Anemonia sulcata green fluorescent protein asFP499
AF322222	Anemonia sulcata nonfluorescent red protein asCP562
AF246709	Anemonia sulcata GFP-like chromoprotein FP595
AF168419	DsRed Discosoma sp. fluorescent protein FP583
AF168420	Discosoma striata fluorescent protein FP483
AF168421	Anemonia majano fluorescent protein FP486
AF168422	Zoanthus sp. fluorescent protein FP506
AF168423	Zoanthus sp. fluorescent protein FP538
AF168424	Clavularia sp. fluorescent protein FP484

```
P42212    avGFP    ------------------------------------MSKGEELFTGVVPILVELDGDV 22
AY015996  rmGFP    ------------------------------------MSKQILKNTCLQEVMSYKVNLEGIV 25
AF372525  rrGFP    -----------------------------------MD---LAKLGLKEVMPTKINLEGLV 22
AF168419  dsRed    -----------------------------------MRSSKNVIKEFMRFKVRMEGTV 22
AF322222  asCP562  ----------------------------------MASFLKKTMPFKTTIEGTV 19
AF168422  asCP562  -----------------------------------MAQSKHGLTKEMTMKYRMEGCV 22
AF246709  asFP595  ----------------------------------MASFLKKTMPFKTTIEGTV 19
AF322221  asFP499  ----------------------------------MYPSIKETMRVQLSMEGSV 19
AF384683  mcGFP    ----------------------------------MSVIKPIMEIKLRMQGVV 18
AF401282  mfGFP    ----------------------------------MSVIKPDMKIKLRMEGAV 18
AF168424  csFP484  MKCKFVFCLSFLVLAITNANIFLRNEADLEEKTLRIPKALTTMGVIKPDMKIKLKMEGNV 60
AF168420  dsFP483  ----------------------------------MSCSKSVIKEEMLIDLHLEGTF 22
AY015995  spGFP    ----------------------------------MNRNVLKNTGLKEIMSAKASVEGIV 25
AF168423  zsFP538  ----------------------------------MAHSKHGLKEEMTMKYHMEGCV 22
AF168421  amFP486  ----------------------------------MALSNKFIGDDMKMTYHMDGCV 22
AY013824  amGFPxm  ----------------------------------MSKGEELFTGIVPVLIELDGDV 22
                                                       :   :     ::* .

P42212    avGFP    NGHKFSVSGEGEGDATYGKL--TLKFICTTG-KLPVPWPTLVTTFSYGVQCFSRYPDHMK 79
AY015996  rmGFP    NNHVFTMEGCGKGNILFGNQ--LVQIRVTKGAPLPFAFDIVSPAFQYGNRTFTKYPNDIS 83
AF372525  rrGFP    GDHAFSMEGVGEGNILEGTQ--EVKISVTKGAPLPFAFDIVSVAFSYGNRAYTGYPEEIS 80
AF168419  dsRED    NGHEFEIEGEGEGRPYEGHN--TVKLKVTKGGPLPFAWDILSPQFQYGSKVYVKHPADIP 80
AF322222  asCP562  NGHYFKCTGKGEGNPFEGTQ--EMKIEVIEGGPLPFAFHILSTSCMYGSKTFIKYVSGIP 77
AF168422  asCP562  DGHKFVITGEGIGYPFKGKQ--AINLCVVEGGPLPFAEDILSAAFNYGNRVFTEYPQDIV 80
AF246709  asFP595  NGHYFKCTGKGEGNPFEGTQ--EMKIEVIEGGPLPFAFHILSTSCMYGSKTFIKYVSGIP 77
AF322221  asFP499  NYHAFKCTGKGEGKPYEGTQ--SLNITITEGGPLPFAFDILSHAFQYGIKVFAKYPKEIP 77
AF384683  mcGFP    NGHKFVIKGEGEGKPFEGTQ--TINLTVKEGAPLPFAYDILTSAFQYGNRVFTKYPDDIP 76
AF401282  mfGFP    NGHKFVIEGDGKGKPFEGTQ--SMDLTVKEGAPLPFAYDILTTVFDYGNRVFAKYPQDIP 76
AF168424  csFP484  NGHAFVIEGEGEGKPYDGTH--TLNLEVKEGAPLPFSYDILSNAFQYGNRALTKYPDDIA 118
AF168420  dsFP483  NGHYFEIKGKGKGQPNEGTN--TVTLEVTKGGPLPFGWHILCPQFQYGNKAFVHHPDNIH 80
AY015995  spGFP    NNHVFSMEGFGKGNVLFGNQ--LMQIRVTKGGPLPFAFDIVSIAFQYGNRTFTKYPDDIA 83
AF168423  zsFP538  NGHKFVITGEGIGYPFKGKQ--TINLCVIEGGPLPFSEDILSAGFKYGDRIFTEYPQDIV 80
AF168421  amFP486  NGHYFTVKGEGNGKPYEGTQTSTFKVTMANGGPLAFSFDILSTVFKYGNRCFTAYPTSMP 82
AY013824  amGFPXM HGHKFSVRGEGEGDADYGKL--EIKFICTTG-KLPVPWPTLVTTFSYGIQCFARYPEHMK 79
                   * *   * * *    *     . .    * *..    :      ** :    :   :
```

```
P42212     avGFP    QHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEV--KFEG---DTLVNRIELKGIDFKEDG
134
AY015996 rmGFP    --DYFIQSFPAGFMYERTLRYEDGGLVEIRSDI--NLIE---DKFVYRVEYKGSNFPDDG 136
AF372525 rrGFP    --DYFLQSFPEGFTYERNIRYQDGGTAIVKSDI--SLED---GKFIVNVDFKAKDLRRMG 133
AF168419 dsRED    --DYKKLSFPEGFKWERVMNFEDGGVVTVTQDS--SLQD---GCFIYKVKFIGVNFPSDG 133
AF322222 asCP562  --DYFKQSFPEGFTWERTTTYEDGGFLTAHQDT--SLDG---DCLVYKVKILGNNFPADG 130
AF168422 asCP562  --DYFKNSCPAGYTWDRSFLFEDGAVCICNADITVSVEEN---CMYHESKFYGVNFPADG 135
AF246709 asFP595  --DYFKQSFPEGFTWERTTTYEDGGFLTAHQDT--SLDG---DCLVYKVKILGNNFPADG 130
AF322221 asFP499  --DFFKQSLPGGFSWERVSTYEDGGVLSATQET--SLQG---DCIICKVKVLGTNFPANG 130
AF384683 mcGFP    --DYFKQTFPEGYSWERIMAYEDQSICTATSDI--KMEG---DCFIYEIQFHGVNFPPNG 129
AF401282 mfGFP    --DYFKQTFPEGYSWERSMTYEDQGICVATNDI--TLMKGVDDCFVYKIRFDGVNFPANG 132
AF168424 csFP484  --DYFKQSFPEGYSWERTMTFEDKGIVKVKSDI--SMEE---DSFIYEIRFDGMNFPPNG 171
AF168420 dsFP483  --DYLKLSFPEGYTWERSMHFEDGGLCCITNDI--SLTG---NCFYYDIKFTGLNFPPNG 133
AY015995 spGFP    --DYFVQSFPAGFFYERNLRFEDGAIVDIRSDI--SLED---DKFHYKVEYRGNGFPSNG 136
AF168423 zsFP538  --DYFKNSCPAGYTWGRSFLFEDGAVCICNVDITVSVKEN---CIYHKSIFNGMNFPADG 135
AF168421 amFP486  --DYFKQAFPDGMSYERTFTYEDGGVATASWEI--SLKGN---CFEHKSTFHGVNFPADG 135
AY013824 amGFPXM MNDFFKSAMPEGYIQERTIFFQDDGKYKTRGEV--KFEG---DTLVNRIELKGMDFKEDG 134
                   *:   : * *    *   ::* .      :   ..       :        . .:   *

P42212     avGFP    NILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQN--TPIGDGP 192
AY015996 rmGFP    PVM-QKTILGIEPSFEAMYMN--NGVLVGEVILVYKLNSGKYYSCHMKTL---MKSKGVV 190
AF372525 rrGFP    PVM-QQDIVGMQPSYESMYTN--VTSVIGECIIAFKLQTGKHFTYHMRTV---YKSKKPV 187
AF168419 dsRED    PVM-QKKTMGWEASTERLYPR--DGVLKGEIHKALKLKDGGHYLVEFKSI---YMAKKP- 186
AF322222 asCP562  P-------------------------------------------------RDAEQS- 137
AF168422 asCP562  PVM-KKMTDNWEPSCEKIIPVPKQGILKGDVSMYLLLKDGGRLRCQFDTV---YKAKSVP 191
AF246709 asFP595  PVM-QNKAGRWEPATEIVYEV--DGVLRGQSLMALKCPGGRHLTCHLHTTYRSKKPASA- 186
AF322221 asFP499  PVM-QKKTCGWEPSTETVIPR--DGGLLLRDTPALMLADGGHLSCFMETT---YKSKKE- 183
AF384683 mcGFP    PVM-QKKTLKWEPSTEKMYVR--DGVLKGDVNMALLLEGGGHYRCDFRST---YKAKKR- 182
AF401282 mfGFP    PVM-QKKTLKWEPSTEKMYVR--DGVLKGDVNMALLLEGGGHYRCDFKTT---YKAKKF- 185
AF168424 csFP484  PVM-QKKTLKWEPSTEIMYVR--DGVLVGDISHSLLLEGGGHYRCDFKSI---YKAKKV- 224
AF168420 dsFP483  PVV-QKKTTGWEPSTERLYPR--DGVLIGDIHHALTVEGGGHYACDIKTV---YRAKKAA 187
AY015995 spGFP    PVM-QKAILGMEPSFEVVYMN--SGVLVGEVDLVYKLESGNYYSCHMKTF---YRSKGGV 190
AF168423 zsFP538  PVM-KKMTTNWEASCEKIMPVPKQGILKGDVSMYLLLKDGGRYRCQFDTV---YKAKSVP 191
AF168421 amFP486  PVM-AKKTTGWDPSFEKMTVC--DGILKGDVTAFLMLQGGGNYRCQFHTS---YKTK-KP 188
AY013824 amGFPXM NILGHKLEYNFNSHNVYIMPDKANNGLKVNFKIRHNIEGGGVQLADHYQTN--VPLGDGP 192


P42212     avGFP    VLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK---- 238
AY015996 rmGFP    KEFPSYHFIQHRLEKTYV-EDGG-FVEQHETAIAQMTSIGKPLGSLHEWV 238
AF372525 rrGFP    ETMPLYHFIQHRLVKTNV-DTASGYVVQHETAIAAHSTIKKIEGSLP--- 233
AF168419 dsRED    VQLPGYYYVDSKLDITSH-NEDYTIVEQYERTEGRHHLFL---------- 225
AF322222 asCP562  RKMG----ASHRDTL------------------------------------ 148
AF168422 asCP562  RKMPDWHFIQHKLTREDRSDAKNQKWHLTEHAIASGSALP---------- 231
AF246709 asFP595  LKMPGFHFEDHRIEIMEE-VEKGKCYKQYEAAVGRYCDAAPSKLGHN--- 232
AF322221 asFP499  VKLPELHFHHLRMEKLNI-SDDWKTVEQHESVVASYS-QVPSKLGHN--- 228
AF384683 mcGFP    VQLPDYHFVDHRIEILSH-DNDYNTVKLSEDAEARYSMLPSQAK------ 225
AF401282 mfGFP    VQLPDYHFVDHRIEILSH-DKDYNKVKLYEHAEA-HSGLPRQAK------ 227
AF168424 csFP484  VKLPDYHFVDHRIEILNH-DKDYNKVTLYENAVARYSLLPSQA------- 266
AF168420 dsFP483  LKMPGYHYVDTKLVIWNN-DKEFMKVEEHEIAVARHHPFYEPKKDK---- 232
AY015995 spGFP    KEFPEYHFIHHRLEKTYV-EEGS-FVEQHETAIAQLTTIGKPLGSLHEWV 238
AF168423 zsFP538  SKMPEWHFIQHKLLREDRSDAKNQKWQLTEHAIAFPSALA---------- 231
AF168421 amFP486  VTMPPNHVVEHRIARTDLDKGGNS-VQLTEHAVAHITSVVPF-------- 229
AY013824 amGFPXM VLIPINHYLSTQTAISKDRNETRDHMVFLEFFSACGHTHGMDELYK---- 238
                    :         :
```

SEQ ID 1 Amino acid sequence of GFP

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFF
KDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

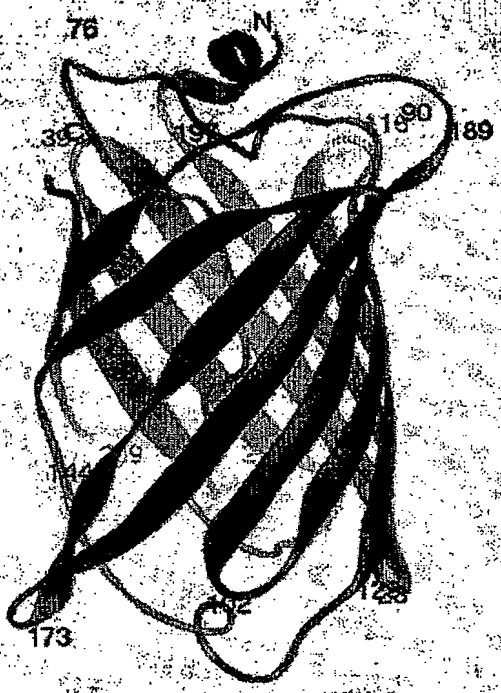
SEQ ID 2 Amino acid sequence of GFP Y66W

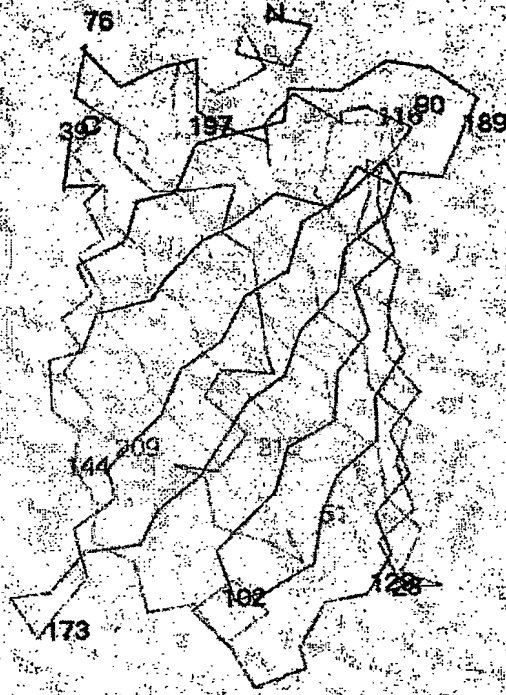
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTFSWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFF
KDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

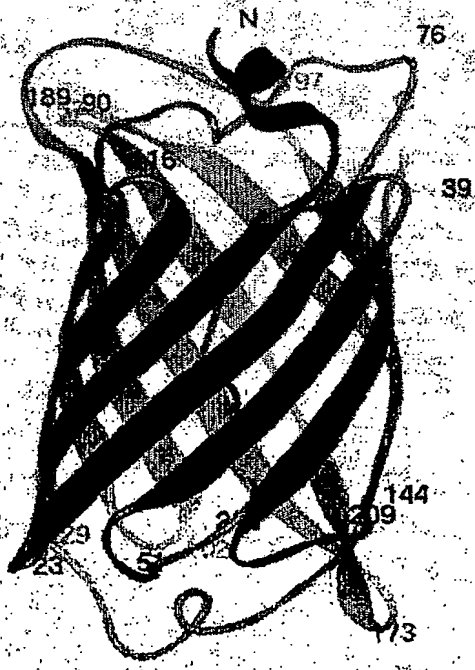
SEQ ID 3 Amino acid sequence of GFP Y66H

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTFSHGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFF
KDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

Figure # 1







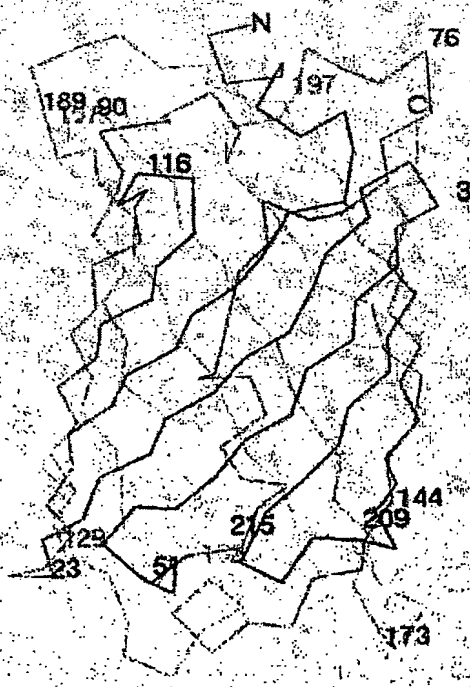


Fig. 1

Positions of appropriate fluorescent protein splitting sites are shown on ribbon and wire frame representations of GFP. The two representations show the same sites from two sides (molecule rotated approximately 180 degrees around a vertical axis).